

FFPMR 101-11.4
April 28, 1972

DATA AUTOMATION PROGRAM RECORDS - GENERAL RECORDS SCHEDULE NO. 20

Introduction

This schedule covers machine readable records, related documentation required for their servicing, and files related to the automatic data processing (ADP) procurement, operations, and management functions.

The decision table format, rather than the columnar format, is used for two reasons: 1) footnote requirements are greatly reduced with this format as compared to the columnar format of the first 19 schedules, and 2) the number of times a given file of logical records has been processed is often more important than the name assigned to it. For example, in an update system, the last created version of an interim master file becomes a final master file after the sponsor declares it error free. The only difference between it and its predecessors is the version number. There may be many versions of a given file created during the course of a processing cycle. Failure to promptly return unneeded tapes to the inventory will lead to excessive requirements for tape. For this reason it is imperative that the creator of machine readable records assign file retention times at the outset--that is to say, at the time of the original system design effort.

The principal machine readable and supporting records common to more than one agency have been divided into four categories. These classes of records correspond roughly to the typical organizational and functional structure found in most ADP installations and their parent organizations.

Data automation planning and operational records (part I) are normally those created during the life cycle of individual computer installations. They deal with planning for, managing, procuring, selecting, utilizing, and accounting for the physical facility investment of the ADP installation and supporting activities.

Documentation required for servicing machine readable records (part II) is defined as the organized series of descriptive documents required to initiate, develop, operate, and maintain specific applications of ADP systems. These include project documentation, system specifications, test data and procedures, file and user documentation, and the various installation procedures and standards used in daily operations.

Erasable media (part III) covers all devices which store machine readable records in an erasable mode. At present, only magnetic media are commonly used for such purposes. However, future technological developments may provide the same characteristics (nonvolatility and easy reusability) now found on magnetic tape.

FPMR 101-11.4
April 28, 1972

Since magnetic records may be destroyed by overwriting, a variety of protective devices and techniques have been developed over the years to preclude inadvertant erasure of records. The earliest technique, still in use, consists of a mechanical interlock device known as a "write protection ring," inserted or left out of a reel of tape. With the later development of computer-manufacturer-supplied "operating systems," an additional safeguard was inserted into the software. It consists of writing file identification and expiration dates on a label record at the beginning of each reel of tape. Other magnetic media, such as disks, depend almost solely on such software devices.

Nonerasable media (part IV) covers such media as punched cards and paper tape. Most ADP installations use media other than magnetic for a variety of roles and functions, but for the most part they are temporary. However, punched cards are sometimes used as documents, such as checks, savings bonds, and requisition forms. In such cases, the functional retention period, developed in other records schedules, will apply.

Procedural analysis of data processing systems (part V) is a guide for archivists, records officers, and auditors in determining secondary uses for data files. Unlike paper, computers create more working copies, which should be erased promptly. But the secondary value, such as furnishing data for audit trails and statistical analysis must be recognized when appraising machine readable records. Many systems, in becoming more automated upon procurement of newer ADP equipment, drop certain manual controls. Since many systems are dynamic, they change due to corresponding changes in legislation and other factors. Thus, nonoperational programs may have to be kept for site-audit records.

Part I. Data Automation Planning and Operational Records

Covering documentation relating to objectives, concepts, policies, and plans providing overall aspects of data automation data needs and systems design of management supporting systems and operational supporting systems, including equipment selection and statistics.

File designation	Consisting of	Which are	Then
1 Planning documents	master plan, feasibility studies with associated charts and diagrams, and supporting data that reflect on the characteristics of the data automation activity	graphic, narrative, and tabular information relating to the present and/or planned ADP composition and requirements of the data automation activity	disposal not authorized by this schedule.
2 Program management	development of plans, policy, and procedures governing the conversion to electrical machine operations and the supervision, control, coordination, and operation of the mechanization program	maintained at policy determination level	disposal not authorized by this schedule.
3 Hardware selection	agency requirements, specifications for hardware, software, and support capabilities of vendors of complete installations or of major peripheral equipment	selection criteria for procurements in the establishment or modification of an ADP installation	dispose of 2 years after specific configuration of equipment is discontinued.
4 Standardization	data elements and codes, standardization requests, and justification for all data systems	promulgated Federal or national (except record copies at National Bureau of Standards)	dispose of when superseded or obsolete.
5		other standards; e.g., developed by agency	disposal not authorized by this schedule.
6 Utilization and maintenance	forms or cards that equipment operators complete relative to machine use, nonuse, or maintenance	used for daily management of operations	dispose of after 3 years.
7	daily detail cards, intermediate summary decks, related magnetic files, and machine listings		dispose of after 90 days
8	monthly summary of cost and utilization reports	card decks, magnetic tape files, and machine listings	dispose of after 3 years.
9 Accountability	documents concerning the management of ADPE equipment	original records maintained at data-processing installation	dispose of 2 years after the date equipment is discontinued.
10	requirements for cards, paper and magnetic tape reels, and inventory of ADPE supplies		dispose of after 1 year.

Part I. Data Automation Planning and Operational Records

File designation	Consisting of	which are	Then
11	contractor's invoices for rental and other charges incurred for use of ADPE		dispose of after 3 years.
12 Magnetic tape library control records	library transaction records	cards decks and magnetic tape files	dispose of when the fourth update cycle is created.
		machine listings	dispose of after 90 days.
		transaction slips	dispose of after 90 days or when no longer needed.

FPMR 101-11.4
April 28, 1972

PART I

NOTES:

Item 13. Machine listings of library transactions are often produced daily. Quite often the transaction listings provide audit trails of the last recording made on a specific reel and may be useful in retrieving a lost file or in determining how a file may have been inadvertently scratched. Accordingly, some installations keep some copies of these listings for as long as 1 year.

Item 14. Transaction slips for military-classified or other sensitive records have longer retention periods. These retention periods are generally specified as a matter of agency policy or regulation.

Part II. Documentation Required for Servicing Machine Readable Records

Documentation covering the organized series of descriptive documents relating to all aspects of system development and operation. These include system planning documents, ADP systems specifications, application program manuals, systems operating instructions, and various management aids.

File designation	Consisting of	Which are	Then
1 Specific data systems planning records	documents containing definition of the system including the system objectives, request for the system, authorizing directives, source data, detailed studies reflecting advantages and disadvantages of alternate solutions, equipment requirements, tangible benefits, output requirements, and schedule for completion	at departmental level headquarters	disposal not authorized by this schedule. review after 5 years.
2		supplementary files at ADP unit level	dispose of 5 years after final action.
3 System test documentation	system test specifications, test runs, machine listings of test data, and test results	an approved system	dispose of 1 year after discontinuance of the system.
4		a disapproved proposed system	dispose of 1 year after final action.
5 Systems design specifications	documents containing operating procedures for implementation of a specific data system, including policies, instructions, details of computer technique, logic charts, and input/output document flow data	for systems for which related magnetic tape data is authorized for blanking	dispose of at time final magnetic tape records produced by system have been blanked.
6		for systems for which the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape.
7 Files(s) specifications	narrative description of the source and functional characteristics of the file(s), a definition of the content of each record in terms of the relative position name, length, and type of each data element in a field (run layout) explanation of the coding system and a cross reference code manual of every code used together with all their values	for systems for which the related magnetic tape data is authorized for blanking	dispose of at time final magnetic tape records produced by system have been blanked.
8		for systems for which the related magnetic tape data is not authorized for blanking	retain with the related magnetic tape.

Part II. Documentation Required for Servicing Machine Readable Records

File designation	Consisting of	Which are	Then	
9	Input specifications	detailed description of each transaction that generated some activity in the system in the form they appear at the time they enter the computer system; identification title, recording media, purpose, frequency, volume, and source; detailed description of the contents of each input to the basic record file and a graphic illustration of each	for systems for which the related magnetic tape data is authorized for blanking	dispose of at time final magnetic tape records produced by system have been scratched.
10			retain with the related magnetic tape.	
11	Output (report forms) specifications	detailed descriptions of products of the system that are to be used outside the computer center	a listing of the outputs by sequence, name, media, purpose, frequency, volume and distribution; a detailed record description; and samples of output in the form of layouts or copies, keyed to names and numbers in the output listings	dispose of on termination of system by either obsolescence, update, or discontinuance.
12	Application program manual	documents reflecting the latest information for a general description of the function, use, and methodology of the program	a description of input, files, and output; source and object code listings and flow diagrams showing the logic of the program; description of program output messages; and coding information, test plan, program test, and operating instructions	dispose of on termination of system by either obsolescence, update, or discontinuance.
13	User guides	information used in training or explaining overall system	handbooks, guides to data availability, and procedures for querying files	retain with systems specification.
14	System operating procedures	user oriented instructions: 1) to prepare input data, 2) for control and interpretation of output reports, and 3) for processing work on the computer	for systems for which the related magnetic tape data is authorized for disposition	dispose of at the time magnetic tape reels are scratched.
15			retain with file (systems) specifications.	

FPMR 101-11.4
April 28, 1972

Part II. Documentation Required for Servicing Machine Readable Records

File designation	Consisting of	Which are	Then
16 Report	printed final report containing the statistical tabulation and an analysis of the findings of a study or survey including a narrative description of methodology employed	for systems which require retention of related magnetic tape data	retain one copy of the printed report with related file specifications.

Part III. Erasable Media

The term erasable media refers to tape (analog, digital), drums, disks, disk packs, data cells, and other devices that store data in an erasable mode. The term "dispose of" in column 4 is synonymous with the terms "scratch," "erase," and "blank."

File designation	Consisting of	Which are	Then
1 Scratch tape (blank tape)	temporary magnetic tape used by the console operators or tape handlers to facilitate general computer runs such as sort and merge runs	new tape or tape not included in a tape library control or files whose retention dates have expired	available for immediate use or reuse.
2 Test tape	magnetic tape used in testing a proposed system	used by programmer for individual run testing and not under library control	dispose of after system has been accepted or discontinued, whichever is sooner.
3		system debugging test data	retain until related program is discontinued.
4		system acceptance test data	
5 Program tape or disk pack	tapes (disk packs) containing sequence of instructions required to accomplish the processing of data or solving a problem	updated	dispose of after third update cycle.
6		the last update of specific LDP application used in a terminated system	dispose of after agency has exhausted its use of the tape.
7		required in audit trail	dispose of in accordance with functional guidelines provided by GAO.
8 Raw data input	magnetic tapes containing data abstracted from source documents or other media and entered into the system for the first time	used for updating with existing program and required to support reconstruction of master file	dispose of first generation data upon successful completion of fourth processing machine pass.
9		not required to support reconstruction of master file and/or used as input for a one-time study or survey	dispose of after raw data is processed into final data and proved satisfactory.
10		officially designated to replace or serve as the basic source data in lieu of the "hard copy" or other input source document	dispose of in accordance with instructions applicable to the "hard copy" or other files documenting the same process, transaction, or case.

Part III. Erasable Media

File designation	Consisting of	Which are	Then
11 Working tape (input/output)	magnetic tape containing output or control within or from one run to a subsequent run that manipulates, sorts and/or moves data through the systems; includes checkpoint, edit, correction, reject list, unmatched data eliminating error, and rerun tapes	used in an updated system	dispose of after subsequent magnetic tapes that contain the accepted detail data have been created and proved satisfactory.
12		used in a one-time study or survey	dispose of after master data tape has been proved to be satisfactory.
13 Valid transaction	magnetic tapes containing valid file of items used with a master data tape input file for creation of master data tape output file	partially valid transaction after all outstanding items are liquidated from current status tapes	dispose of after third update cycle.
14		valid transaction after cumulative final master tape is prepared and determined to be successful, and there is no necessity for statistical analysis	dispose of after third update cycle.
15		used in additional statistical analysis	disposal not authorized by this schedule.
16 Information retrieval system master reference	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file (including the security copy tape of data on disk packs)	a cumulative index to scientific and technical publications, and bibliographic and other nonrecord material	dispose of after third update cycle.
17		an index to record material such as correspondence, legal hearings and decisions, patents and trademarks, and record copy of publications	disposal not authorized by this schedule.
18 Federal loan and grant program master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file (initial data includes excerpts from forms placed in case files)	cumulative data of funds made available through federally supported loan and grant programs	dispose of after third update cycle.
19		noncumulative periodic file of status of Federal loan and grant activity	disposal not authorized by this schedule.

Part III. Erasable Media

File designation	Consisting of	Which are	Then
20 "Housekeeping systems" master data file	magnetic media containing data for such "housekeeping systems" as fiscal accountability, supply management, and payroll administration	not required for GAO site audit	dispose of in accordance with instructions applicable to the hard copy or other files documenting the same process, transaction, or case.
21		required for GAO site audit	dispose of in accordance with functional guidelines provided by GAO.
22 Economic statistics master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data such as status of banks and insurance institutions; production, consumption, and monetary status of industry and agriculture; value of foreign commerce and other economic indicators such as construction of houses and buildings; motor, rail, and air travel; communications, including broadcasting, telephone, and telegraph	
23		noncumulative data used to prepare reports covering a limited period of time	disposal not authorized by this schedule.
24		noncumulative recurring periodic surveys including wholesale and consumer price indexes, annual industry, housing vacancy, and other economic indicators	disposal not authorized by this schedule.
25		noncumulative economic census taken during 5-year intervals	disposal not authorized by this schedule.
26 Social statistics master file	magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative social and demographic data concerning births, deaths, and marriages; income taxes paid; social security accounts; employment information; law enforcement, crime and civil disturbance, and other social indicators	dispose of after third update cycle.
27		noncumulative data used to prepare reports covering a limited period of time	disposal not authorized by this schedule.

FPMR 101-11.4
April 28, 1972

Part III. Erasable Media

File designation	Consisting of	Which are	Then
28		noncumulative recurring periodic surveys including current population statistics, annual industry, housing vacancy, voter participation, and statistics of income sample	disposal not authorized by this schedule.
29		noncumulative demographic censuses	disposal not authorized by this schedule.
30 Natural resources master file	continuously updated magnetic media containing data created by the merging of prior master file with valid transaction data to create a new master file	cumulative data on characteristics, use, and ownership of natural resources such as land, water, minerals, and timber	dispose of after third update cycle.
31		noncumulative data used to prepare reports covering a limited period of time	disposal not authorized by this schedule.
32 Longitudinal studies master data file	magnetic tape containing data recorded over time from one or more sources	a series of observations relating to individual units (persons, places, things)	disposal not authorized by this schedule.
33 Scientific data files	magnetic media source data recordings received from experimental sensor instruments for scientific measurements such as outer space orbiting spacecraft, oceanographic and geophysical phenomena, and medical research (including analog tape)	converted to raw data digital magnetic tape media	dispose of after meaningful data has been analyzed.
34		not converted or converted only in part to raw data digital magnetic tape media	dispose of after determination has been made that the data will not be converted to raw data digital magnetic tape media.
35	magnetic media containing data created either from analog magnetic tape or recorded directly on magnetic digital tape for scientific measurements of astronomic, outer space, and oceanographic phenomena; air and water quality; and medical research measurements	held in national data centers	disposal not authorized by this schedule.
36		not duplicated in national data centers	disposal not authorized by this schedule.
37		duplicated in national data centers	dispose of after determination is made that data is not required outside the data centers.

Part III. Erasable Media

File designation	Consisting of	Which are	Then
38		not calibrated or validated	dispose of after subsequent magnetic tapes containing the accepted data have been created and proved satisfactory.
39	Summary data file	magnetic tape containing aggregates of individual observations from valid transaction or master data files that are disaggregates of published data	substantially unpublished, such as tapes containing data that are disclosure free
40	Publication tape	magnetic tape containing source output data extracted from the system (without destroying the source tapes)	reproduced and disseminated as a publication or used for reproducing a printed publication
41	Print tape		used for producing required printouts of tabulations, ledgers, tables, registers, and reports
42	Re-formatted data file	magnetic tape containing essentially duplicate data from the master data file but which is created for use with other computer hardware systems	created for the specific purpose of information interchange
43			of specific application for agency computer hardware systems
44	Sample and subsample data files	magnetic tape containing individual observations selected from a larger census or survey file such as stratified or pure random sample files with or without weighting factors	disclosure free or useful in statistical analysis or policy formulation models and simulation studies
45	Security backup file	magnetic tape that is identical in format to master tape retained as security in case master tape is damaged or inadvertently erased	updated
46			a one-time study or survey
47	Other agency files	magnetic tape created by other agencies	not altered substantially by the receiving agency
			dispose of when no longer needed.

FPMR 101-11.4
April 28, 1972

PART III

NOTES:

Items 3 and 4. This type of data is differentiated from simple debugging test data in that the data set is used to exercise all possible data system options within the complete set of programs. System debugging test data means data used to debug individual programs or groups of programs prior to final acceptance testing. It must be retained until the related program is discontinued.

Acceptance test data may also be a contractually defined specification or item in software systems being procured and it or a listing of it may have to be kept with the contract file. For details in this case, see General Records Schedule 3, item 4.

In other cases, particularly in systems where accounting for funds is involved, it may be required that the files be kept until a particular version of a system has been audited and approved by the General Accounting Office. Retention periods in this case will be in accordance with the specific functional file in one of the other general records schedules. This means that specific acceptance test data sets might have to be kept for the life of the particular version of a software system or until all records produced under that system have been disposed of.

Item 7. Just as the acceptance test data may need to be kept beyond its useful life for auditing purposes, programs which processed that data may also be kept for audit purposes beyond the operational life of the particular system. Disk packs are relatively expensive for long-term storage and there is usually a backup copy of the system on magnetic tape. In these cases, the tape copy of the program together with all relevant documentation may be used in lieu of the disk pack version. Either source or object versions of the system may be used for this purpose.

Items 16, 18, 22, 26, and 30. "Cumulative date" implies no earlier data is deleted in the present pass.

Part IV. Nonerasable Media

Nonerasable media refers to ADP punched cards, paper tape, and other nonerasable, machine readable media.

File designation	Consisting of	Which are	Then
1	ADP program card files	punched cards containing common language source program data (source deck)	processed with a processor or utility program to produce a machine-coded object program
2		machine-punched cards containing coded machine language instructions arranged in proper sequence (object deck)	read into computer memory before running a program to cause the computer to perform data-processing functions
3		prepunched utility or processor program card decks	used to update installation systems software
4		job stream (job stack, job control) card decks	used to activate program-processing modules performing a data-processing job

Part IV. Nonerasable Media

File designation	Consisting of	Which are	Then
5	ADP program control cards	punched cards containing data for program control generated by the producer or user	pertinent to a specific run or cycle dispose of individual cards or sets of cards when replaced by new cards and when necessary changes (if any) have been made to appropriate data-processing manual.
6			for repetitive use and updated either by ADP or user dispose of individual cards after replacement by new cards; destroy control deck 1 year after program has been removed from system, or after system has been discontinued.
7	ADP source data cards (or paper tape as applicable)	punched cards or paper tape containing data abstracted from source documents and used for conversion to magnetic tape or processing on (EAM) electric accounting machine equipment created after January 1, 1970	retained by ADP operational elements as backup to magnetic tape or disk dispose of when related magnetic file has been proven to be satisfactory and has grandfather backup.
8		EAM output listings and reports	dispose of after 180 days if used in processing without being converted to magnetic tape.
9		on magnetic tape	dispose of after verification of data on related magnetic tape.
10		punched cards that contain original entry data with film or written inserts	source documents dispose of in accordance with instructions applicable to the hard copy or other files documenting the same process, transaction, or case.

PART IV

NOTES:

Items 5 and 6. These items refer to parameter cards associated with the execution of various options of operational programs. These include date cards, periodic (monthly or quarterly) options executed only occasionally, and queries to information retrieval systems. They do not include card decks for generalized interpreter systems used with computer simulation software packages such as SIMSCRIPT, GPSS, DYNAMO, and similar systems. These decks have the status of program source decks. Similarly, all except report generation decks in file management systems are considered to be source program decks and should be retained or destroyed in accordance with the criteria of items 5-7 of part III.

Approved For Release 2001/08/30 : CIA-RDP74-00005R000100010007-6

April 28, 1972

Part V. Procedural Analysis of Data Processing Systems--Guidelines for Appraising Files and Data Sets for Permanent Retention

1. Introduction

This section is a guide to ADP systems analysts, records officers, and archivists for determining the nature of data files (also called data sets) generated by computers. Factors that influence the selection of specific data files for permanent retention in machine readable form (chiefly on magnetic tape) are indicated and explained here.

In examining a variety of documentation files for different ADP systems, substantial differences were found in the use of technical terms among agencies, and in some cases, within agencies. These differences are being resolved by several vocabulary standardization groups, among them Federal Information Processing (FIP) task group 5 and its successors and the American National Standards Institute (ANSI) X.3.5 committee on vocabulary. However, the definitions in the vocabulary have not been standardized to the extent that flow chart symbols have been in ANSI Standard X3.12-1968, Flowchart Symbols and Their Usage in Information Processing.

Accordingly, better guidance for appraising data and documentation files can usually be achieved by studying the high-level system flow charts in addition to the narrative description found in the system documentation files. The system files are enumerated and described in part II of the schedule. This section has been written based on the fact that virtually all ADP systems are composed of a small number of basic procedure types connected in sequences that can be called modules. The text and charts in the following sections are organized around this concept. Almost all existing ADP systems can be analyzed into portions or groupings of these charts.

2. The Elements of Data-Processing Systems

Data processing systems are composed of four basic classes: hardware, software, peopleware, and data files. The hardware consists of the central processing unit and all of its peripheral devices and recording media. The software consists of the machine instructions that direct the hardware to perform the processing. Peopleware is listed in parts I and II and consists of specifications, hard copy documentation, and user manuals for all personnel involved in running a system. The data files themselves are listed and described in part III. Appraisal criteria for them will constitute the bulk of sections 3 and 4, below.

FPMR 101-11.4
April 28, 1972

2.1 Hardware

Computer hardware and recording media are still undergoing relatively rapid evolution, and this presents a problem in finding equipment that can successfully read some older machine readable files. Files to be retained permanently may have to be recopied periodically onto newer media or totally converted in format and most other physical characteristics. Since costs for this type of work are declining, this situation presents no undue burden to the holder of this data. In general, the property value and conversion costs of machine readable records are less than one-tenth of 1 percent of the data collection and editing costs of the information recorded on it. Upon consultation, the Office of the National Archives, National Archives and Records Service, will recommend procedures and techniques needed for the physical preservation of the record content beyond the life of the recording medium.

2.2 Software

This is divided into two main types, systems and application software. System software is furnished by the computer manufacturer and is designed primarily to manage the available resources of the computer complex in an efficient manner. The computer complex consists of the central processor and its attached peripheral devices, such as card readers, magnetic tape drives, high speed printers, and other equipment. In general, this type of software is not related to any specific file or record maintained in an installation. It is, therefore, of no permanent value except to the history of the development of computer science. Selected portions of systems software specifications are useful for reading files produced on one computer with another equipment configuration. However, this information may be documented in less than one page and does not require extensive documentation. Subclasses of system software include utility, operating system, sorts, merge, and compiler software.

An exception is application software written in one of the standardized machine independent programming languages. COBOL, FORTRAN, and PL-1 are the three most widely used of such languages. In most cases, application software written in these languages may be considered for retention with the related files. However, only a small portion of the total software written for an application need be retained permanently. For example, a file that has been closed off and covers a specific period of time will not be updated. Therefore, the update software is unlikely to be ever required again and is disposable.

FPMR 101-11.4

April 28, 1972

If the file is a large complex data base designed to service many inquiries, retention of the query software may be warranted. However, much good query software is commercially available to handle the problems of file inquiry. Therefore, retention of this software is less important than retention of user documentation described in section 2.3 below.

The final class of software that may have permanent value is that used in computer simulation work. There are several software systems that have been used in policy formulation and evaluation work for high-level management in agencies. The three best known such software systems are SIMSCRIPT, DYNAMO, and GPSS (General Purpose Systems Simulator). Like COBOL and FORTRAN, these systems are available for most computers on the market. It is also likely that they will continue to be available for the foreseeable future. What is important to save in such applications are the source program decks. The policy alternatives and much of the information on a project is contained in these decks and they often constitute records of intrinsic historical value. Economic and financial projection models and war game software are typical examples.

2.3 Peopleware

A wide variety of hard copy documentation is produced in data processing systems. Peopleware is that documentation required by the personnel involved in the design, development, operation and maintenance of ADP systems. The files are listed in parts I and II of this schedule. Of interest in this section are primarily those files required for the direct servicing of files declared permanent.

The basic concept to grasp in data processing is that the record constitutes a representation of an event and not the event itself. As such, the representation or record may have been recorded by a sensor (as in scientific measurement) or may have been transcribed and encoded from some other document or document group as in all transaction reporting. In either case, a researcher needs to know what kind of transformations occurred between the actual event and its representation on magnetic tape. This knowledge is in the documentation described in part II of this schedule.

For example, most housekeeping systems usually encode events using elaborate code tables rather than narrative fields on the record. A payroll system may have dozens of deduction code possibilities as well as an equal number of pay plans. Typical codes would represent bond deductions, local tax rates for States and municipalities, bond and charity deductions, overtime and premium shift differential rates, etc. In scientific work, instrument readings represent observations of physical phenomena and other occurrences.

Each time a transaction is encoded or instrument reading is made, there is a possibility of an error or distortion taking place in the process. The errors may be simple random occurrences such as digit transposition by key-punch operators or transcribers, or systematic because of some bias in the recording instrument or observer. In general, the scientist attempts

FFMR 101-11.4
April 28, 1972

to calibrate his instruments and adjusts instrument readings for other known factors, and the systems accountant devises consistency checks, batch totals, and clerical training programs to assure "accurate" recording of his data.

Permanent records of this class include the file and input specifications (items 7-10 of part II) along with the final version of the related tape file. They tell a future user of the probable quality and coverage of the file and, for those with much encoding such as accounting files, the meaning of all of the descriptive data fields along with the bias and judgement that went into transcribing a record of an event into a coded element. Some portions of system-operating procedures and user guides (items 13-15 of part II) are also useful for later reference work. These records are essential for determining how the related data files were used for operations and research and must be retained even if the related software is disposed.

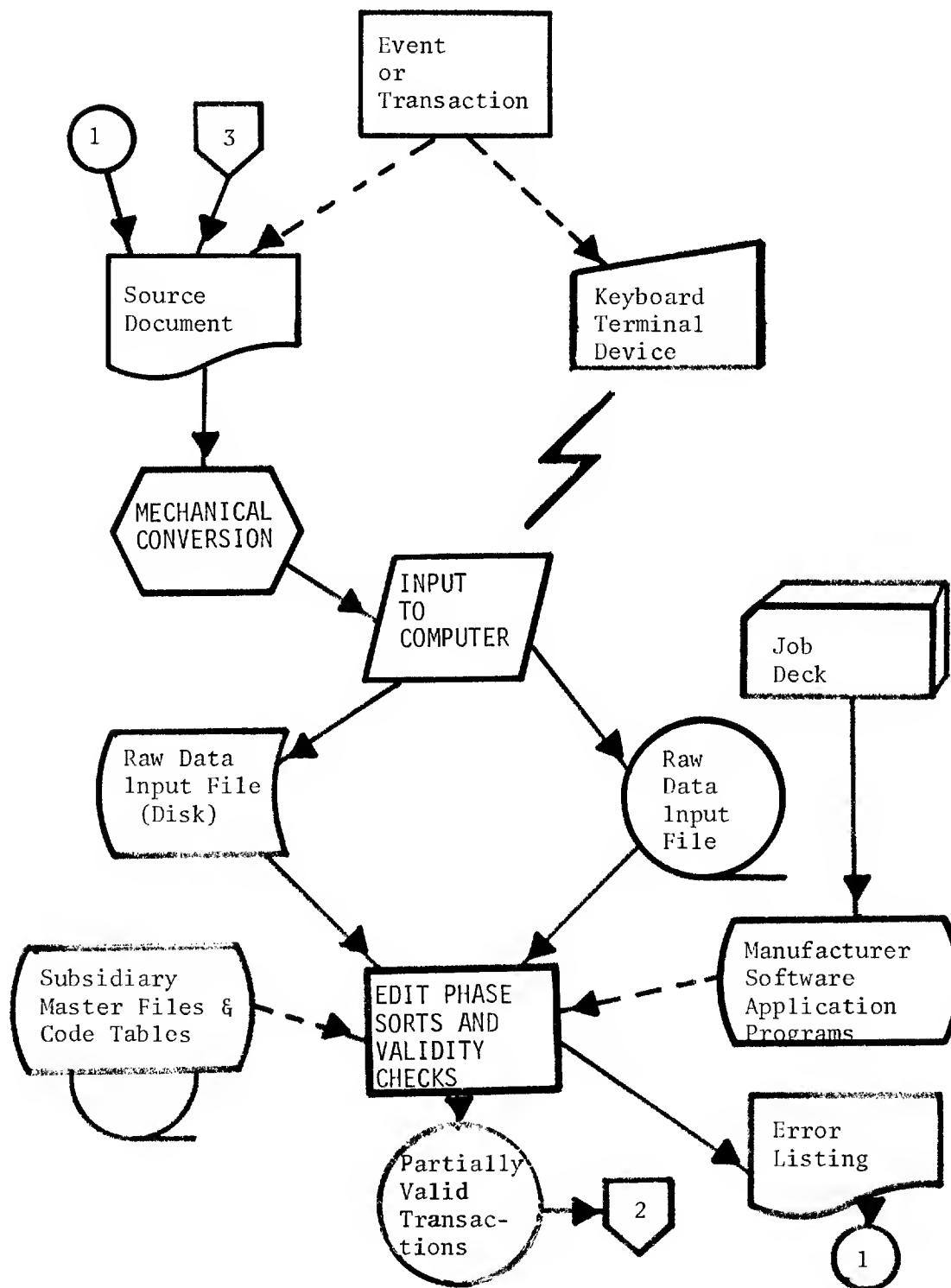
3. Data-Processing Systems Flow Charts and Their Use in File Appraisal

Data records in ADP systems are processed both manually and mechanically before finally residing in a file as a correct record. This section presents typical systems charts found in the high-level documentation of most such applications. These charts should be used by an appraiser for determining which files among many are most useful for permanent retention.

Data processing systems are categorized by two sets of terms. One breakdown is between continuing and one-time systems; the second is between real-time and batch-processing systems. Real-time systems handle one transaction at a time and complete the function of posting and validation before going on to the next transaction. These operations occur at the time the actual real world event occurs or at the latest, soon afterward. Batch-processing systems perform one stage of processing for a group (or batch) of transactions. These operations occur after the real world event took place. The delays may range from hours in some cases to months in others.

Continuing systems are those which are run periodically with a repetition rate ranging from a few hours to a year. Most housekeeping systems are of this type. The most familiar applications are for payrolls, inventory control, and financial management. Although the file contents are continually changing, such systems have high continuity from one period to the next and are well documented for auditing and operational purposes. One-time systems are less well documented than continuing systems. There is usually pressure to deliver the results in the form of reports within a time constraint. Many undocumented ad hoc decisions are made during the course of these projects to meet project deadlines, which result in files containing data errors that may not reflect the contents of published results. Surveys, simulation projects, and censuses fall into this category.

GENERALIZED INPUT UPDATE (1)



Various combinations of these attributes are found in ADP systems. Real-time systems are used in continuous housekeeping systems, the most notable being military command and control and airline reservation applications. Batch processing is used for both one-time and continuous systems and constitutes the bulk of existing data-processing applications.

The files in all types of systems are increasingly being put on mass storage devices such as disks rather than tape. However, even in real-time systems, backup and recovery procedures dictate that magnetic tape copies of the file be created. These are usually called either "file dumps" or "safe data dumps." They are also created for running off summary reports since total file scanning of disk files is inefficient. These files can be appraised in the same manner as tape resident files.

3.1 Input and Update Subsystem Phases

These two phases are common to all data-processing systems that involve file maintenance. Two typical flow charts are shown and labelled "Generalized Input Update." They show the processing steps taken to record, convert, check, edit, and post a record to a file for later use.

3.1.1 Source Data Conversion Phase

Data can be converted to machine readable form by several methods. Formerly, data were transcribed from source documents onto transcript sheets. They were keypunched, converted to magnetic tape, and then processed. More recent methods either record machine readable data onto source documents (turnaround documents) or accept input directly into computers through keyboard-driven terminals (source data automation).

The data control function is closely interwoven with the mechanical conversion process. One part of data control consists of keeping count of the documents in each batch to be processed and control the totals of one or more quantity fields. Examples would be dollar totals and counts of checks or invoices. The other part of data control is the manual editing of source documents. Such editing consists of checking codes and resolving errors.

The typical sequence is shown on the chart entitled "Generalized Input Update I." A manual-handling phase is followed by a media conversion step. The machine readable transaction is further validated by a series of computer runs. Errors may be introduced and detected during each stage of the process. Correction procedures depend on the stage at which the error is detected, the type of error, and the conversion hardware used. For example, in keypunch-oriented systems, verification is used to minimize conversion errors while the computer passes are used to catch logical and transcription errors. Some systems combine error detection and correction processes by attaching the conversion equipment to a computer.

FPMR 101-11.4

April 28, 1972

However many steps and cycles occur in the process, the end product is usually called a raw data tape. Other names used are unsorted transactions, partially edited transactions, sorted preliminary update file, and the like.

Raw data tapes are seldom of any permanent value since they contain erroneous and duplicate records. A possible exception in the case of real-time systems where the tape may be named "logging file" in such applications as message switching or production control systems. Usually such tapes are kept for a short period of time as backup to recreate a real-time file. The sole other usage is for system test data or some simple transaction counting for real-time system work load studies.

3.1.2 Edit Phase-Sorts and Validity Checks

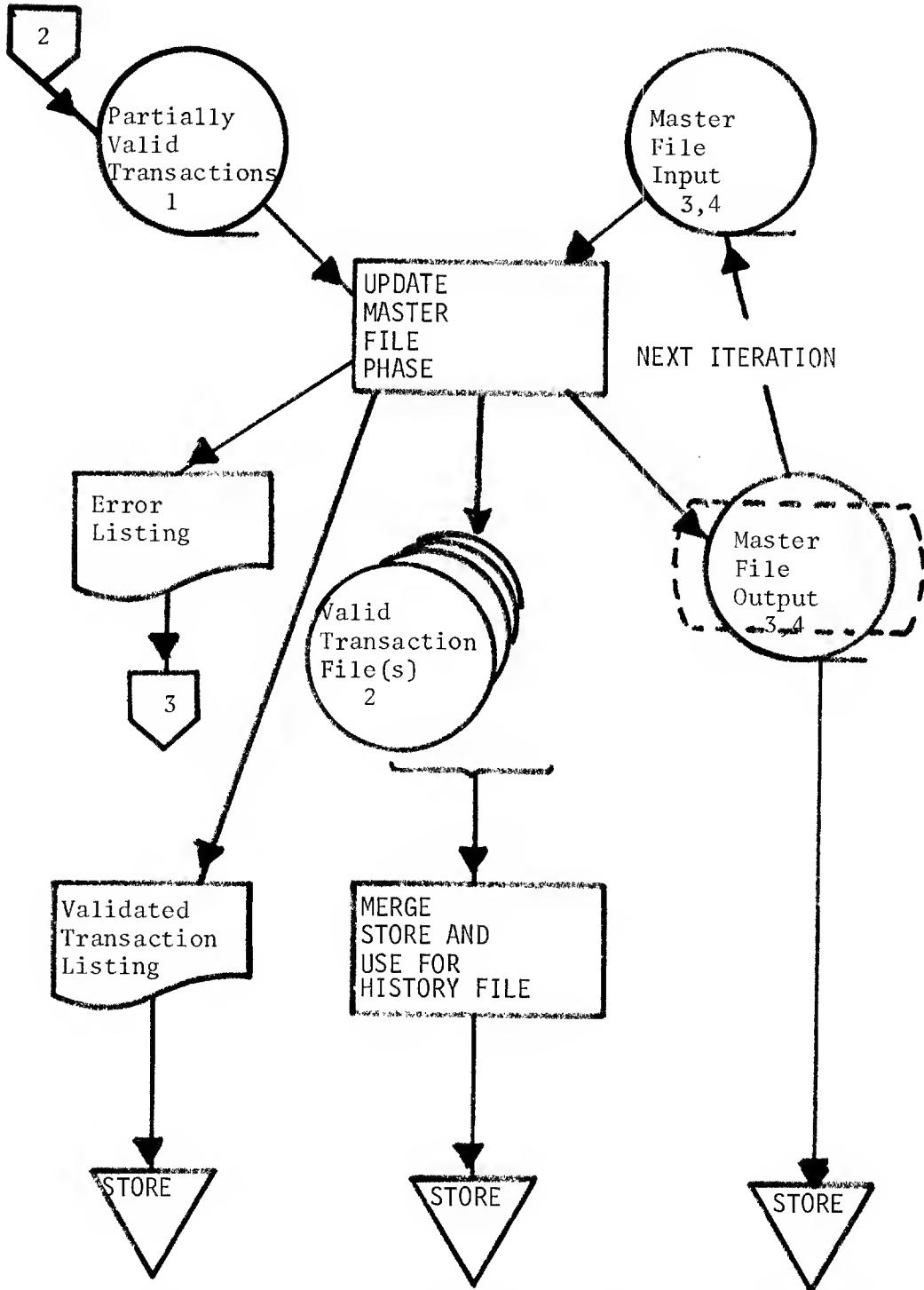
Many tapes with records of temporary value are produced at this stage. Another common designation for these tapes is work tapes. This phase or module processes transaction files against various editing and validation criteria. These criteria may be found in a computer program, such as a table of valid transaction codes, or in a subsidiary master file, such as valid account numbers or a name and address file. Other common checks performed here are for numeric characters in quantity fields, transaction batch totals, transaction counts, and consistency checks.

The output of such a phase is a file of partially validated transactions. Two methods are used in handling errors at this time. In one invalid transaction items are listed on a printer along with the error indicators for immediate correction on a batch basis. In other systems, erroneous records are coded to indicate the presence of certain errors, but are not deleted from the transaction file. Instead, they are kept in the transaction file for still further checks in the update program itself. This gives a consolidated error listing at one time for a given batch of transactions. The most common additional tests performed on the data would be tests against the master file key itself. Examples are transactions that attempt to delete a nonexistent record or insert a duplicate record into a file; others may be quantities that are checked for "reasonableness."

3.1.3 Update Master File or Data Base Phase

The edit phase's output is the updating run's input. This is shown in "Generalized Input Update II." If the file contains only one application, such as accounts payable or receivable, it is normally called a "master file." If it contains data from a series of applications, or summary data from a variety of sources, then it is a "data base." Under certain circumstances, an individual transaction and status report is the direct concern of an organization's top management. Examples are status reports for an important research and development project, construction job, or loan--all of which might have enduring value.

GENERALIZED INPUT UPDATE (II)



1 or work tape.

2 A series of these files may be merged into a continuous history file.

3. May be cumulative or noncumulative file.

4. Until final version is approved, prior versions are interim master files.

April 28, 1972

There are several types of master files. In some systems, they include only currently active records. Purged master file records are periodically transferred to dormant account (history) files. Personnel and payroll files are typical, with the purge generally occurring at year's end. Many other files are cumulative and continue to grow in size depending upon the application. As in the case of paper records, frequency of use is the major criterion when deciding the length of time "detailed transactions" will remain on the master file.

In the case of periodically updated files, where transactions are deleted, purged records are often merged into historical files. These files are valuable. However, they may lack data found on the master record file. Therefore, both the merged periodic transaction file and the master file should be retained. Items 13-15 of part III denote the types of transaction files created. Items 16-32 of part III furnish disposition criteria for master files.

Master files are seldom updated for a given period in a single update pass. Some errors cannot be detected until the actual posting attempt is made. This creates a series of "interim" master files. The only valid file would be the one from which the periodic output was run. Usually, processing deadlines determine which version is "valid." Interim master files are usually retained for short periods as backup tapes for the final master file. (See items 44 and 45 of part III.) This retention plan is called the "grandfather system." In appraising master files for permanent retention, it is preferable to retain the "as of" date from the official files. While this is usually possible, there are many cycle-billing systems in which the master file is never completely purged of detail transactions--thus never complete. In such cases, it may be preferable to retain extracts of the master file made for reporting purposes and audit trails rather than the master file itself. Items 39-43 of part III describe alternative selections to master record tapes.

3.2 Report Generation Phase

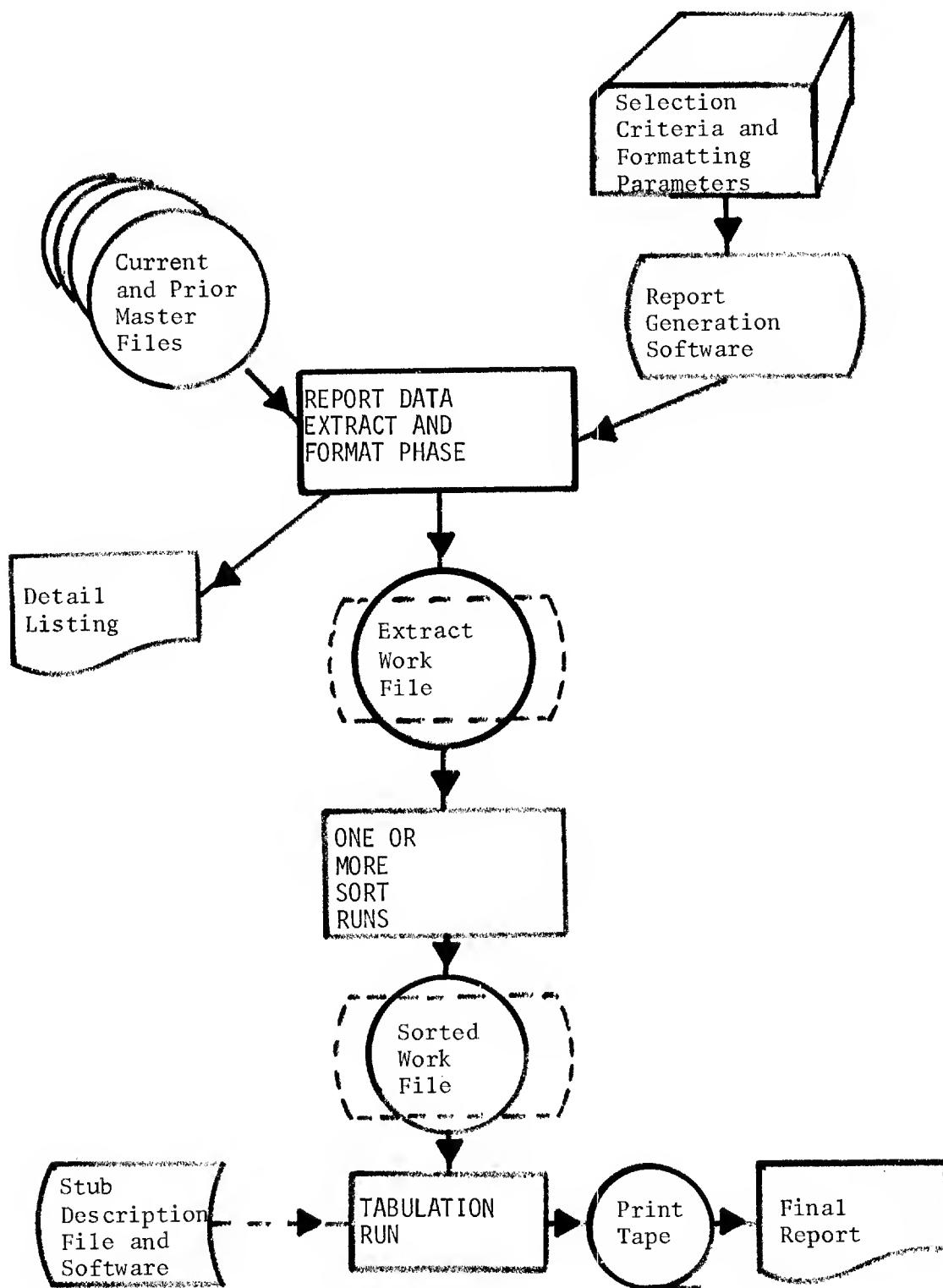
This section describes the files, processing, and software used to produce output from ADP systems. The chart labeled "Report Generation Phase" shows the usual processing sequence in such modules from the machine readable record to the final printed report or listing. Since mass, random-access storage is increasingly used, the chart shows tape and disk files used interchangeably, although in practice one or the other medium will predominate.

As indicated previously, unless its usefulness or transferability to future computing systems is assured, it is unnecessary to retain output-oriented software. This evaluation should be ad hoc.

3.2.1 Report Data Extract and Format Phase

If the printed report and the master file are in identical sequence, the data selection, tabulation, and printing phase may occur in one program.

REPORT GENERATION PHASE



This is characteristic of billing, payroll, and most housekeeping systems. However, often it is necessary to print the report in a sequence different from that of the master file. The use of a sort will resequence the selected records as desired, creating a series of intermediate work files between the master file and the printed output. The flowchart labeled "Report Generation Phase" shows both tapes and disk files in the processing sequence.

Newer computers and operating systems seldom produce work tapes except for the largest multireel files. The intermediate files reside on disk as transient files within the "job stream." (See item 4, part IV.) The input file, the job control deck, and the final printed output only are visible to the uninformed. Thus, seldom is there need to retain intermediate files because they can always be recreated from the master file.

In one case, extract files are useful and should be retained. Files that contain "statistical samples" of the entire body of data often have long-term value when the methodology is documented. These sample files, along with appropriate weighting factors and stripped of identifying information disclosing individual persons or establishments are immediately releasable for public research.

The next problem is to determine which of several work files to retain. In general, this depends upon the degree of decoding stub descriptors required to interpret the file. Heavily encoded files with little or no narrative description are suitable provided that the stub descriptor files and tables required for human reading and interpretation are of reasonable length. When the code is a Federal Information Processing Standard, the length of the code table is unimportant. An example is the table of State and county codes of the United States with more than 3,000 entries. Tables of less than 200 entries developed for individual agencies or one-time studies may be reasonably left encoded as they can be decoded by simple computer programs. For large code tables stub descriptions are preferable for long-term preservation.

3.22 One or More Sorting Runs

Extract files are often in the wrong sequence for producing reports or required tabulations. In fact, the same file may be sorted into as many as 10 different sequences for different types of analysis and tabulation. The criteria for retention of sorted work tapes are the same as for extract tapes described above. The output of this phase is a sorted work tape or file ready for tabulation, summarization, and editing.

This type of file on tape is often a useful research file, particularly if there has been some editing and if interpreting has been performed. In general, continuing administrative systems have relatively few processing steps between the first extract run and the final output pass. This is different from one-time reports as described in section 4 below.

3.2.3 Tabulation Run

The inputs to the tabulation run are the sorted work files, and usually with a stub descriptor file. The stub descriptor file is invariably used when a very large list of codes must be displayed in plain text. If it is on magnetic tape, one or more sorting runs are typically required of the extract tape in order to apply the descriptors. If on disk, most of this decoding can be performed during the tabulation run. When files are considered for retention, the information necessary to decide such data elements must be retained. This may either be a hard copy document as described in section 2.3 above or a machine readable file.

The final output of a tabulation run may consist of either summary data files or a print tape (items 39 and 41 of part III). Summary data files may also serve as publication tapes (item 40 of part III) when they are reproduced and disseminated to the public and/or Federal agencies. Summary data files are occasionally used as input for published and widely disseminated printed reports.

Many installations do not use print tapes when they produce a computer listing or report. Instead, the data are temporarily transcribed on disks until the printed report is complete. However, tapes can be created upon specific request if there is a known demand and further use for the same information in machine readable form. This procedure is often followed by producers of general proposal statistics.

When a tape file may be classified under more than one of these three categories, disposal is not authorized by this schedule.

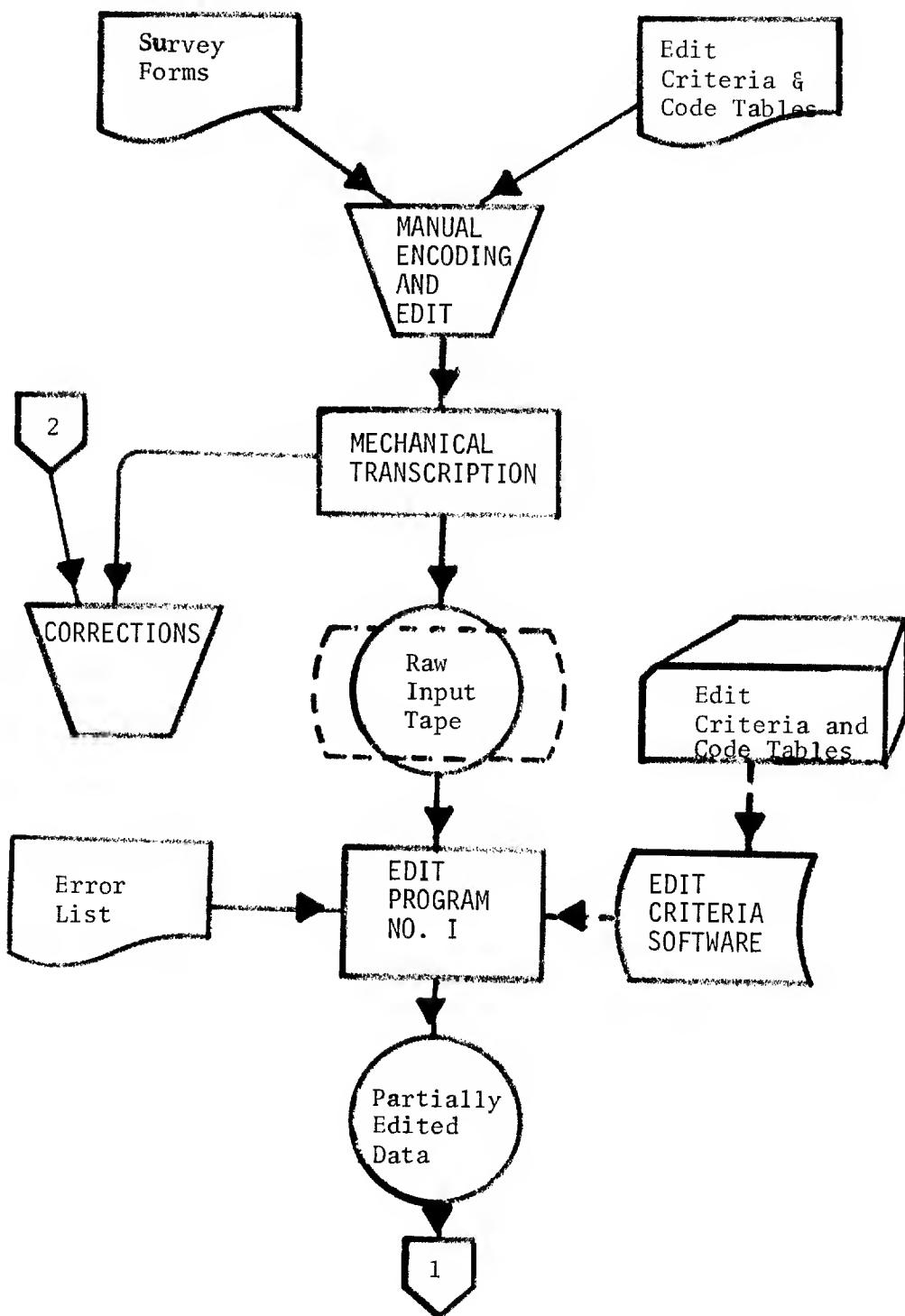
4. One-Time Surveys and Report Generation Systems

The sequence of operations in one-time surveys, censuses, and tabulations is shown in the following two charts. When the flow process charts are compared to the typical continuous file maintenance system, the similarities are evident. The basic difference between continuously running systems and one-time jobs is the much higher amount of manual editing and encoding required. Unless the job is a very large effort with many thousands of observations, the forms used allow somewhat more variability in field entries than accounting type documents. Since line respondents to these surveys rarely have an opportunity to correct the inputs, much more manual editing and encoding is required to correct (clean up) a file prior to its use in tabulations.

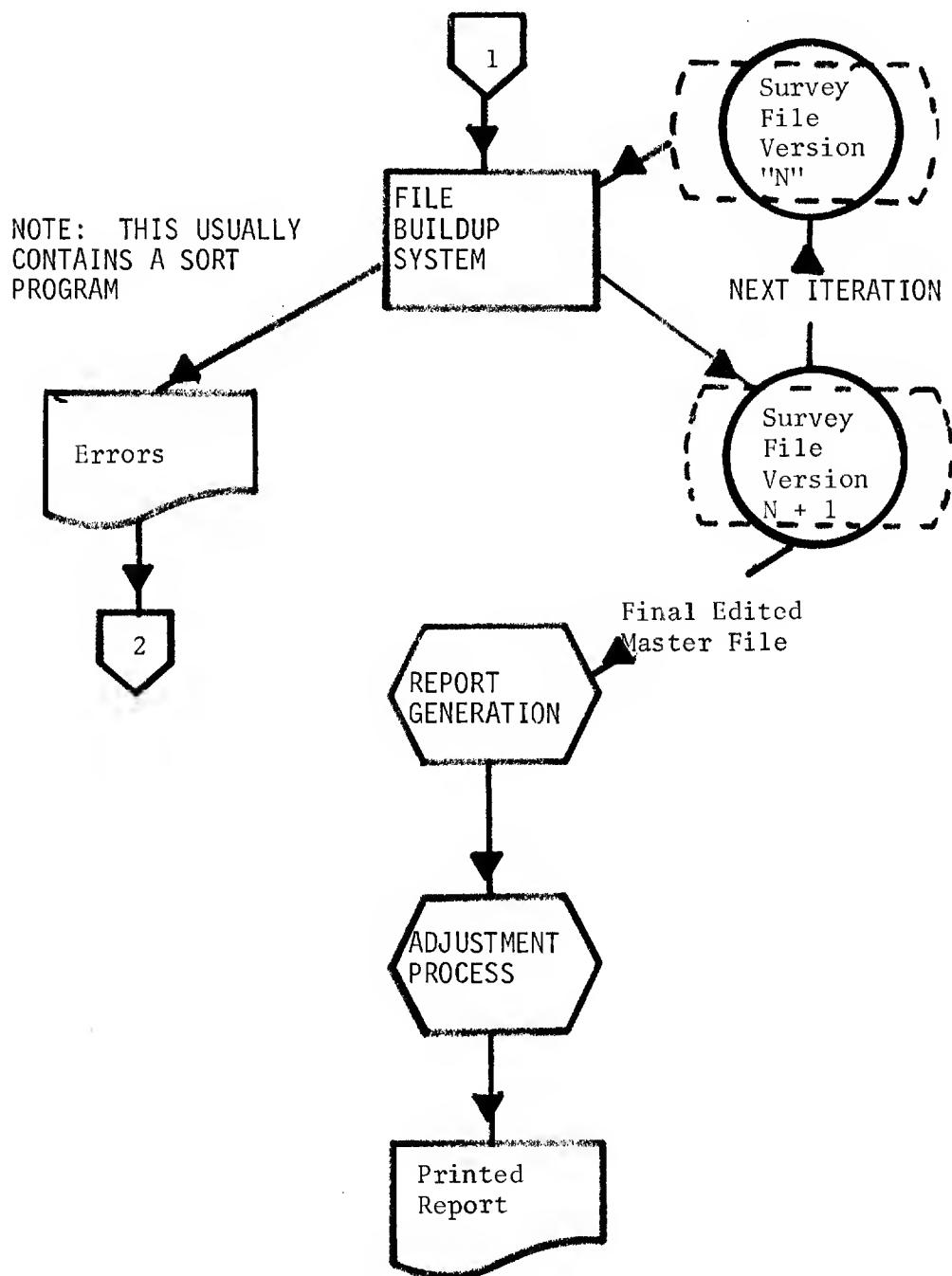
As the second sheet shows, there is a file buildup process which occurs with no changes occurring to individual records after they have entered the file. Where a multiplicity of systems and sources feed the file, the individual records are usually of variable length to minimize storage requirements.

Documentation for such systems contains complex record formats but includes few of the elaborate codes found in administrative systems.

TYPICAL ONE-TIME PROCESSING SEQUENCE (1)



TYPICAL ONE-TIME PROCESSING SEQUENCE (11)



FPMR 101-11.4

April 28, 1972

When these files are retained, it is important to identify the original sources of information, the instructions to respondents for filling out the forms, together with sample forms, and the directions given to the response form editors for proper interpretation and secondary usage of these files. Most such files are described in part II of this schedule.

5. Adjustments in One-Time Jobs

A variety of additions, changes, and deletions can occur in either individual or groups of entries within a file at any stage of processing. They can occur for a variety of reasons and lead to magnetic tape files of different accessibility and validity. If the files are to be retained permanently, it is important to document the corrections and adjustments.

This record of changes constitutes the equivalent of the accountant's audit trail for evaluating the accuracy of a financial file.

If a payroll record contains an error, the originating office usually hears about it in short order, particularly when an employee is short changed. In sample surveys, respondents seldom correct reporting and transcription errors unless elaborate procedures have been established for a review of the machine-prepared record. This correction and review process almost always occurs in accounting systems, while cost usually precludes this process in most one-time jobs.

Therefore, such files of recorded observations contain a variety of errors which in summary tabulations are nonsense entries. For example, male widows may appear in a tabulation. Such errors arise from a number of causes. Correction of the tabulations can be made at any step in the process between the final survey file and the printed report. The accompanying flow chart shows the points in the procedure where this is usually made.

If the error is thought to be a random event, the illogical counts are generally distributed to all other possible categories and deleted from the tabulation array. This would lead to a discrepancy between the published table and the final master file. The illogical records would remain in the file uncorrected.

Systematic final edited master errors also occur frequently in encoding and processing. In these cases, the summary file may be corrected by moving the entire nonsense count to the correct location in the table. These errors can also be corrected in the final master file using the computer.

Another common adjustment operation occurs when a tabulation discloses individual confidential information. Confidentiality is protected in one of three ways: 1) by deleting the entry on the summary file and combining with enough other tabular entries to eliminate individual disclosure, 2) by correcting the print tape, and 3) by correcting only the printed report. In the first method the summary file is releasable to the public. In the second, the summary file is not releasable to the public, but the print file is.

WHERE OBSERVATIONS AND SUMMARIES GET ADJUSTED IN ONE-TIME JOBS

